SUPPLEMENTARY DISCUSSION

Further insight into the mechanism of transcription blockage could be obtained by applying the strategy developed by Grabczyk and Usdin (1), which is based on competition between different transcription DNA templates present in the same reaction mixture during multiple-round transcription experiment.

The rate of transcription on a DNA template depends upon both the rate of transcription elongation, and the rate of transcription initiation; the latter depends upon an apparent initiation rate constant, and the concentration of free RNAP in solution. The elongation rate and an apparent initiation rate constant are intrinsic properties of a given template, while the concentration of free RNAP depends upon all templates present in solution, because all templates temporarily sequester RNAP during their transcription, thus decreasing concentration of free RNAP; and the template with a lower elongation rate (for example, due to reversible pausing at certain sequences) sequesters RNAP for a longer time, thus more strongly decreasing free RNAP concentration in solution than the template with higher elongation rate.

In the case of high excess of RNAP over DNA templates, when most of the RNAP molecules remain free in solution, different DNA templates are transcribing independently and the yield of their transcripts is defined only by intrinsic properties of the respective templates. In particular, at sufficiently high RNAP concentrations to provide frequent transcription initiation, the yields would depend upon rates of elongation.

In the opposite case, when the concentration of DNA is high enough that most of the RNAP molecules are bound to DNA (and thus the concentration of free RNAP is low) the initiation step becomes rate-limiting; consequently, if the templates have the same apparent initiation rate constant, the overall rate of transcription would be the same for all templates regardless of their intrinsic elongation rates.

This "equalizing" of transcription rates in mixture of the plasmid containing long GAA/CTT repeats and parental control plasmid without repeats (while in separate transcription reactions the former has significantly lower transcription rate than the latter) was observed by Grabczyk and Usdin (1), and allowed them to conclude that the decreasing transcription rates for these repeats is due to RNAP sequestration.

However, this equalizing wouldn't occur if one of the templates had a greater probability of irreversible transcription interruptions (for example, increased probability for RNAP to "fall off" the template during transcription). That is because, after irreversible transcription interruption, a new initiation event would be needed to resume the transcription; thus, more initiation events would be required to obtain one full-size transcription product. Consequently, in the equimolar mixture of plasmids with the same initiation rate constant (which should be the case for our plasmids because they have the identical sequences up to more than 200 bp downstream from the promoter), the plasmids with increased probability of irreversible transcription interruption would yield less full-length transcription products regardless of the concentration of RNAP.

That is what was observed in our experiments for the mixture of plasmids with the nick upstream of the G16 insert (which produces strong blockage) and a control plasmid (see Fig.4 and its legends for details): the amount of full-length transcript (run-off) remains

lower for the plasmid with the blockage even at low RNAP concentrations and high DNA concentrations, when templates are competing for RNAP. That suggests that at least some of blockage events are irreversible. Also, under the conditions of competition for RNAP, a decrease in the yield of control plasmid transcription upon addition of G16-containing plasmid with nick (which produces strong blockage) was only slightly less than upon addition of the same amount of G16-containing plasmid without nick (which produces less than 1% blockage under these conditions). That suggested that the blockage is not accompanied by strong sequestration of RNAP.

These experiments also revealed that at high RNAP concentrations and low DNA concentrations the blockage for the G16 plasmid with nick is greater than at low RNAP concentrations and high DNA concentrations (Fig.4). The increase in blockage upon increase in RNAP concentration manifests itself as a decrease in yield of the full-sized transcription product in comparison to that with a plasmid which doesn't produce blockage (Fig.4), as well as an increase of the intensity of the blockage signals relative to the full-size product (see legend to Fig.4). We think that the increased blockage upon increased RNAP concentration could be explained by the contribution from collisions between RNAP molecules (see Discussion, the main text).

SUPPLEMENTARY MATERIALS AND METHODS

Transcription substrate preparation

Plasmids were purified using the standard maxi-prep protocol (Qiagen), except that the denaturation step was reduced to several seconds, and an additional ethanol precipitation

was done following isopropanol precipitation. Nicking reactions with enzymes Nt.BspQI and Nt.BbvCI (NEB) were performed for ~ 3 hours at 44°C and 37°C, respectively, under buffer conditions recommended by the manufacturer, stopped by addition of EDTA up to 20 mM, column purified using nucleotide-removal kit (Qiagen) and analyzed by agarose gel, which verified that plasmids were converted to the open-circular (nicked) form. In the case of linear substrates, either supercoiled or nicked plasmids were digested by HindIII restriction enzyme and gel purified using the Qiagen protocol (except for experiments in Fig.S4, where all substrates were purified using nucleotide removal kit) with following modifications: (i) cutting from the gel was performed using a reference, so that the purified plasmid sample was neither stained by ethidium bromide, nor UV-irradiated, and (ii) dissolving of agarose was performed 30-40 min at room temperature, instead of 10 min at 50°C. The concentration of purified linear DNA was estimated by gel-electrophoresis.

Transcription

The in vitro T7 transcription reaction was performed for 30 min at 37°C in 12 μl of buffer containing 33 mM TrisHCl (pH 7.9), 5 mM $MgCl_2$, 8.3 mM NaCl, 1.7 mM spermidine, 4.2 mM DTT, ~10 ng of DNA substrate, 10 μCi (α-^{32}P)CTP (which corresponds to the final concentration of about 0.0003 mM), 20 units of T7 RNA polymerase and 16 units of RNasin (both from Promega corp, Madison, WI) and various concentrations of unlabeled ("cold") NTPs. Under our standard conditions (designated as A, U, G = 1, C = 0.1) the concentrations were 0.17 mM each for ATP, GTP and UTP, and 0.017 mM CTP; under the conditions (A, U, G, C = 1) the concentration for all NTPs was 0.17 mM; under the

conditions (A, U, G, C = 1/3) the concentration for all NTPs was one-third of this amount, i.e. ~ 0.06 mM). The reaction was stopped by adding 94 μl of buffer containing 1% SDS, 106 mM TrisHCl (pH 7.6), 13.2 mM EDTA, 160 mM NaCl, 25 μg tRNA (Invitrogen) and 10 μg Proteinase K (Invitrogen) and incubated at room temperature for 15 min. Next, 11 μl of 3 M NaOAc (pH 5.27) and 300 μl of 100 % ethanol (cooled at -20°C) were added, the mixture was incubated on dry ice for ~30 min and centrifuged for 20 min at 14,000 rpm at 4°C. The supernatant was removed; then 75 % ethanol (cooled at -20°C) was added, and the mixture was centrifuged for 5 min under the same conditions. The supernatant was removed, and the pellets were dried with a SpeedVac for 10 min and dissolved in 8 μl of the formamide loading solution (94 % formamide, 2 mM EDTA, 0.05 % bromophenol blue, 0.05 % xylene cyanol). Then, 2 μl of the obtained solution were mixed with 2 μl of formamide loading solution, incubated at 85°C for ~2 min, quickly chilled on pre-cooled rack and loaded on a 5% sequencing gel (acrylamide:bis-acrylamide 29:1) containing 8 M urea and run at ~70 V/cm. As size markers, 5'-end labeled denatured DNA ladders consisting of DNA fragments of sizes increasing in steps of 100 or 10 bases were used. Then, the gel was dried and exposed to a phosphoimager screen.

Use of "spiking transcript" to improve transcription product quantitation.

In our transcription protocol, after transcription is stopped, transcription products are purified by ethanol precipitation, which could be accompanied by some loss of RNA products. These losses, together with gel-loading errors, could introduce random variation in amounts of RNA recovered from transcription reactions. These variations don't affect evaluation of ratios of yields of transcription products obtained in the same

reactions (like blockage and run-off signals for the same template), but would introduce random errors in evaluation of products obtained in different reactions. To eliminate the effect of these errors, an additional transcription reaction producing a "spiking transcript" was performed separately in parallel with the other transcription reaction, and after that, equal amounts of spiking transcript were added into other transcription reactions after they were stopped, but before their purification. The requirement for choosing a template for the spiking transcript is that the spiking transcript must not co-migrate with any fraction of interest in the gel. The amount of spiking transcript could be chosen in such a way that its signal is roughly comparable with the signals of interest. For practical purposes, it is convenient to stop the spiking transcription reaction by addition of EDTA (usually up to 10-20mM final concentration) and then to add appropriate amounts to the analysed transcription reactions already mixed with the stop buffer. It is reasonable to assume that the random losses during purification, as well as effects of loading error would be practically the same for each fraction within the same reaction, including the spiking transcript. Thus, normalization of transcription signal to the spiked transcript signal from the same reaction should "cancel" error produced by sample purification and gel loading.

SUPPLEMENTARY FIGURE LEGENDS

Fig. S1. Comparison of the transcription yields at A, U, G = 1, C = 0.1 and A, U, G, C =1 conditions.

In these experiments we used plasmid pWT-C (2) which has the same sequence as the plasmids used in the current work, except for the insert sequence which doesn't produce any detectable effect upon transcription. The plasmid linearized by HindIII (493 nt from the promoter) was used as the test template, and the plasmid linearized by ScaI (1877 nt from the promoter) was used to produce a spiking transcript. Transcription was performed either at A, U, G = 1, C = 0.1 or A, U, G, C =1 conditions for "cold" NTP, using the same amount of either (α-^{32}P)CTP (designated C*) or (α-^{32}P)ATP (designated as A*) for radioactive labeling, which was much smaller than the amounts of cold NTPs (see Supplementary Materials and Methods). The spiking transcript was produced at A, U, G =1;C =0.1; C* conditions, and its equal amounts (1/12 part of the standard reaction) were spiked into each transcription mixture after stopping the transcription reaction. The left panel shows the radio-autograph of the gel, the right panel shows the ratio of run-off signals for A, U, G = 1, C = 0.1 and A, U, G, C =1 conditions for either radioactive CTP (C*) or ATP (A*). All signals were normalized to the spiking transcript in the same lane. The average and deviations calculated for two experiments are presented. For C*, the ratio was 2.81 ± 0.07; for A*, the ratio was 0.291 ± 0.025, i.e. about 10-fold smaller, exactly as predicted. At A, U, G, C =1 conditions, the A* signal is about 2-fold times

smaller, consistent with the fact that this transcript sequence contains 2-fold less A than C.

Fig.S2. Blockage signals coinciding with those produced by a nick in the non-template strand are present in intact negatively supercoiled DNA, but not in intact linear DNA. Plasmid pN-aga-0 (see Materials and Methods, the main text) was used in this experiment. L stands for linear DNA, OC stands for open circular DNA, SC stand for negatively supercoiled DNA. Arrow shows position of the nick (after the nucleotide number 248 from the promoter). Double-headed white block arrows with solid borders show blockage signals which could be seen both in nicked and in supercoiled DNA (where they are much weaker). Some signals (shown by dashed-border block arrows) are too weak to be distinguished in negatively supercoiled DNA.

Fig.S3. Effect of a nick in the non-template strand versus that in the template strand for random DNA sequence. Linearized plasmids pN-TS and pN-NTS were used in these experiments (see Materials and Methods). Nick locations (after the nucleotide number 253 and number 256 from the promoter for the nick in the non-template and the template strand, respectively) are shown by black arrows. For a nick in the non-template strand (NTS) (lane 1), an approximate location of the blockage products is shown by the dotted lines (a thicker line shows the area where they are more pronounced). For a nick in the template (TS) (lane 3), the predominant blockage signal and much smaller additional blockage are shown by large and by small white triangles, respectively. The minor signal

might be due to the encounter of a second polymerase with the one that is already stalled at the nick.

Fig. S4. Effect of reduced NTP concentration upon nick-facilitated blockage. All substrates are the same as in Fig.3.

Fig.S5. Effect of the nick in the non-template strand on transcription blockage by the human telomeric sequence in both orientations. Plasmids pN-aga-hTel-G and pN-aga-hTel-C contain the human telomeric sequence in opposite orientations: $(TTAGGG)_{17}$ and $(CCCTAA)_{17}$, respectively. The rest of the plasmids have the same sequence as other pN-aga-(insert) plasmids including a site for the nicking enzyme Nt.BspQI three nucleotides upstream from the insert (see Materials and Methods in the main text, for the flanking sequences), except they have an extra 30 nt region containing U-less cassette and Nt.BbvCI-site in the area closer to the promoter, which were not used in the current work. The telomeric insert spans from nucleotide 282 to 383 from the promoter. Transcription was performed with circular plasmid substrates, and the substrates without nicks are negatively supercoiled.

Fig.S6. Effect of the nick in the non-template strand on transcription blockage for the human c-Myc sequence in both orientations. Plasmids pN-agac-cMyc-G and pN-agac-cMyc-C contain the human cMyc sequence in opposite orientations: GGGAGGGGCGCTTATGGGGAGGG and CCCTCCCCATAAGCGCCCCTCCC, respectively. The rest of the plasmids have the same sequence as the ones with telomeric

inserts (see legend to Fig. S6) except for an extra C following the AGA sequence localized immediately after the nicking site. Thus, the nick is four nucleotides upstream from the insert, which spans from position 283 to 305 from the promoter. Transcription was performed with circular plasmid substrates, and the substrates without nicks are negatively supercoiled.

Fig.S7. Effect of the nick in the non-template strand on transcription blockage for various sequences. Linearized plasmids of pN-aga-(insert) series (see Materials and Methods, the main text) were used in these experiment at (A, U, G, C = 1) NTP concentration. The sequences of inserts are shown above corresponding lanes, and their positions are shown by thin vertical lines.

Fig.S8. Effect on transcription of upstream and downstream nicks on closely (2 nt) located G16 or C16 inserts.

Circular (**A**, **B**) or linearized (**C**) plasmids pN-tc-G16-cc-N and pN-tc-C16-cc-N were used in these experiments. G16 and C16 inserts (from 250 to 265 nt from the promoter) and nicks position (after 247 nt and after 267 nt from the promoter for the upstream and downstream nicks, respectively) are shown by the thick black line and black arrows, respectively. Note that different radioatographs in this figure have different exposures and cannot be used for comparison of run-off product in different radio-autographs.

Fig.S9. Increase in the distance from the nick decreases the blockage.

Nicked circular plasmids pN-tc-G16-cc-N (lanes 1, 2), pN-tc-C16-cc-N (lanes 3, 4), which have respective inserts spanning from position 250 to 265 from the promoter and nicking enzymes sites localized 2 nt either upstream or downsream from the insert; and pN-(9)-G16-(9)-N (lanes 5, 6), pN-(9)-C16-(9)-N (lanes 7, 8) which have respective inserts spanning from position 257 to 272 from the promoter and nicking enzymes sites localized 9 nt either upstream or downstream from the insert were used in these experiments at (A, U, G, C = 1/3) NTP concentration. The designation "u" stands for upstream nick, "d" stands for downstream nick. Positions of nicks (arrows) and inserts (thick black lines) for the lanes 1-4, and for the lanes 5-8, are shown at the left and at the right of the 10 nt ladder, respectively.

Fig. S10. Free energy calculations for RNA/DNA and DNA/DNA duplexes with sequence compositions used in this work. For calculations, HyTer^{TM} server (Peyret, Saro, SantaLucia, http://ozone3.chem.wayne.edu) was used. Free energy ΔG° for the duplexe formation was calculated for 37°C, 5 mM Mg^{2+}, 10 mM monovalent cations, that is close to the conditions of our transcription reaction. (ΔG° under these conditions is negative for all duplexes, meaning that the duplex formation is energetically favorable. For convenience, we plot -ΔG° instead of ΔG°. (The higher the -ΔG°, the more stable the duplex.) Calculated free energies are plotted against G-content for various non-G base X. As always, when talking about sequence, we mean the sequence of the non-template DNA strand. Thus, "homopurine sequence" means the homopurine non-template strand/homopyrimidine template strand in the duplex. In calculations for the RNA/DNA duplex T was replaced by U. The points on graphs correspond to free energies for

sequences shown in Fig.2, with 0, 34, 50, 63, 69, 75, 81, 84, and 100 % of G. (For X=A, all percentages were studied experimentally; for X=C, only 0, 75, 81, 84; for X=T, only 75, 81, 84). The top left panel corresponds to (-ΔG°) for the RNA/DNA duplex, top right panel corresponds to the DNA/DNA duplex, and the central bottom panel corresponds to the -ΔG° difference between for these two duplexes. For RNA/DNA duplexes, the ranking of duplex stabilities (C > A > T) does not coincide with the ranking of intensities of blockages (A > C ≈ T). Moreover, though at given (G, C)-richness (i.e. percentage of (G+C) in the sequence), homopurine sequences are more stable than mixed or homopyrimidine sequences, the effect of (G,C)-richness on stability predominates: for example, the stability of the RNA/DNA duplex for pure C sequence roughly coincides with the highly G-rich (about 70%) homopurine (i.e. X=A) sequence, and that clearly does not correlate with the trends observed for blockages (Fig. 2). The situation strongly changes, when instead of absolute stabilities of RNA/DNA duplexes, we consider the differences in stabilities of RNA/DNA and DNA/DNA duplexes, further referred to as differential stabilities (ΔΔG in the designations of Roy and Lieber (3)): Differential stabilities for homopurine sequences (X=A) are the largest, starting from the low G+C content (about 20%) up to 100% (which covers with an excess the whole area in which we are able to detect the blockage, i. e. from about 33%(GAA) to 100% (pure G)). The differential stability of the pure C sequence coincides with the homopurine sequence with low (about 30%) percentage of G, for which blockage would be barely detectable. Moreover, for G-content above 50%, the difference between C and T substitutions is much smaller than the difference between either of them and A-substitution. These trends in differential stabilities correlate with the trends observed for the blockage (Fig.2). This

correlation is consistent with the R-loop formation as a source of blockage: if during R-loop nucleation there is a reversible strand exchange between the RNA and non-template DNA strand, then the probability of R-loop formation would depend on the difference in stabilities of RNA/DNA and DNA/DNA duplexes, rather than the absolute stability of RNA/DNA duplex. The importance of difference in stabilities of RNA/DNA and DNA/DNA duplexes for R-loop formation has been previously established by Roy and Lieber (3). In addition, the displaced non-template DNA strand could be transiently sequestered by triplex formation (see Discussion), which would also facilitate R-loop formation. Because triplex formation requires a homopurine sequence, that would also contribute to increased blockage by the sequences with A versus those with C and T substitutions. If the RNA-displacing propensity of the non-template strand is decreased, for example by nicking, the absolute stability of the RNA/DNA hybrid might play an increased role in blockage. That might explain why in the presence of a nick, the C-sequence produces stronger blockages than the random sequence. Also, for the human telomeric sequence, calculations predict a more stable RNA/DNA hybrid in the orientation in which G-containing strand is RNA than for the opposite orientation. That is also consistent with our previous study (4) and our present data (Fig. S5).

Fig.S11. Downstream sequence for transcription templates used in this work. The sequence (non-template strand) is shown starting immediately downstream from various unique short sequences shown in the Materials and Methods section up to the end created

by HindIII digestion, which corresponds to run offs for the linear substrates used in this work. This sequence is the same for all plasmids used in this work.

Fig.S12. Estimation of the initiation to elongation time ratios for various NTP concentrations using transcription templates with two different lengths.

In these experiments we used plasmid pWT-C (2) which has the same sequence as the plasmids used in the current work, except for the insert sequence which doesn't produce any detectable effect upon transcription. To produce transcription templates with various lengths, the plasmid was digested by either BamHI, or HindIII, or ScaI, which sites are localized 234, 493, and 1877 nt from the T7 promoter, respectively. BamHI and ScaI digested plasmids were used as the short and the long templates, respectively (their length ratio is about 8); and HindIII digested plasmid was used to produce spiking transcript. The experiments were performed at two three-fold different NTPs concentrations, designated A, U, G, C =1/3 and A, U, G, C =1 (see Supplementary Materials and Methods). The spiking transcript was produced at A, U, G, C =1 conditions, and its equal amounts were spiked into each transcription mixture after stopping the transcription reaction. The left panel shows the gel radio-autograph. At the right, the diagrams show the quantitation of the results. All signals were normalized to the spiking transcript in the same lane. The average and deviations calculated for two experiments are presented. The top diagram at the right shows the ratios of the radioactive signals for the long and the short signals at two NTP concentration. It is seen that they are practically the same (around 4) for both concentrations. Because transcripts are "body-labeled" with

radioactive cytosine, to convert these signal ratios to the ratios of molar yields of transcripts, one have to divide them by the ratio of cytosines in the long and in the short template (this ratio is about 7).

The bottom diagram shows the ratio of run-off signals for larger and smaller NTP concentrations, measured for the short and the long template, It is seen that within the experimental error, these ratios are practically the same for the short and for the long templates and are around 7-8. Because under both conditions the same amount of radioactive cytosine was used (see Supplementary Materials and Methods and Fig.S1,) radioactive cytosine incorporation is inversely proportional to the "cold" cytosine concentration, thus at three-fold lower concentration of all NTPs it would be three-fold higher. Thus, to convert the ratio of signals in the bottom diagram into the molar ratios, they must be multiplied by three. Thus, the molar yields at higher concentration are 20-25-fold higher than at lower concentration.

Calculations: The yield of transcription product is inversely proportional to the time of one transcription cycle, τ, which is a sum of the transcription initiation time, τ_i, and the transcription elongation time, τ_e.

The initiation time does not depend on the length of the template and should be the same for templates, which have the same promoter and the same sequence in close vicinity from the promoter, while the elongation time is proportional to the length of the template, assuming that the templates are sufficiently homogeneous.

That allows to estimate the ratio of the nucleation and elongation times by comparing the yield of transcription for the DNA templates with various lengths and the same promoter area.

Let L_1 and L_2, and Y_1 and Y_2 be the lengths and the molar yields of transcripts for a shorter and a longer templates, respectively.

Then

$$\frac{Y_1}{Y_2}=\frac{\tau_2}{\tau_1}=\frac{\tau_i+\tau_{e2}}{\tau_i+\tau_{e1}}=\frac{\frac{\tau_i}{\tau_{e1}}+\frac{\tau_{e2}}{\tau_{e1}}}{\frac{\tau_i}{\tau_{e1}}+1}=\frac{\frac{\tau_i}{\tau_{e1}}+\frac{L_2}{L_1}}{\frac{\tau_i}{\tau_{e1}}+1}=1+\frac{\frac{L_2}{L_1}-1}{\frac{\tau_i}{\tau_{e1}}+1} \qquad \text{Eq.1}$$

In this derivation, we took advantage of the proportionality of the elongation time to the length of the template.

From this equation, we can obtain the ratio of the initiation and elongation times for the shorter template:

$$\frac{\tau_i}{\tau_{e1}}=\frac{\frac{L_2}{L_1}-\frac{Y_1}{Y_2}}{\frac{Y_1}{Y_2}-1} \qquad \text{Eq.2}$$

and for the longer template:

$$\frac{\tau_i}{\tau_{e2}}=\frac{L_1}{L_2}\frac{\tau_i}{\tau_{e1}}=\frac{L_1}{L_2}\frac{\frac{L_2}{L_1}-\frac{Y_1}{Y_2}}{\frac{Y_1}{Y_2}-1} \qquad \text{Eq.3}$$

The ratio of molar yields can be calculated from the experimentally measured signals

$$\frac{Y_1}{Y_2} = \frac{\frac{S_1}{C_1}}{\frac{S_2}{C_2}} = \frac{\frac{S_1}{S_2}}{\frac{C_1}{C_2}} \qquad \text{Eq.4}$$

where C is the number of cytosines and S is the radioactive run-off signal for respective templates.

Substituting the experimentally measured ratio

$$\frac{S_1}{S_2} = \frac{1}{\frac{S_2}{S_1}} = \frac{1}{4} \qquad \text{Eq.5}$$

and the ratios known from the sequence of the templates

$$\frac{L_2}{L_1} = 8; \quad \frac{C_2}{C_1} = 7 \qquad \text{Eq.6}$$

into above equations, we obtain

$$\frac{\tau_i}{\tau_{e1}} = 8; \quad \frac{\tau_i}{\tau_{e2}} = 1; \qquad \text{Eq.7}$$

Thus, according to this simplified analysis, for the shorter and for the longer templates, the initiation time comprises 8/9 = 90% and 1/2=50% of the transcription cycle time, respectively. We think, that actual contribution of the initiation time to the transcription cycle could be even larger than the one calculated above, because these calculations don't take into account the probability of spontaneous transcription termination during elongation, which would decrease the apparent yield of the run-off product. This decrease would be stronger for a longer template than for a shorter template; thus, the measured ratio for apparent molar yields of shorter and longer templates Y_1/Y_2 would be larger than the same ration in the absence of spontaneous termination. From Eqs.1-3 it is seen that the ratio of initiation and elongation times decreases when the ratio of the transcription yields for shorter and longer templates increases. Thus, we think that our calculations underestimate contribution of initiation time to the time of the transcription cycle.

In any case, at least for the shorter template, the contribution from the initiation time overwhelmingly predominates the time of the transcription cycle for both low and high NTP concentrations.

Consequently, 20-fold decrease of the transcription yield (i.e. 20-fold increase in the transcription cycle time) for the shorter template upon three-fold decrease in NTPs concentration indicating that the initiation time (which according to estimations above provides a major contribution to the transcription cycle time under both NTP concentrations) also should increase about 20-fold upon three-fold decrease in NTPs concentration.

From our experiments it is difficult to make similar reliable estimation for the behavior of the elongation time upon decrease in NTP concentrations, because our calculations

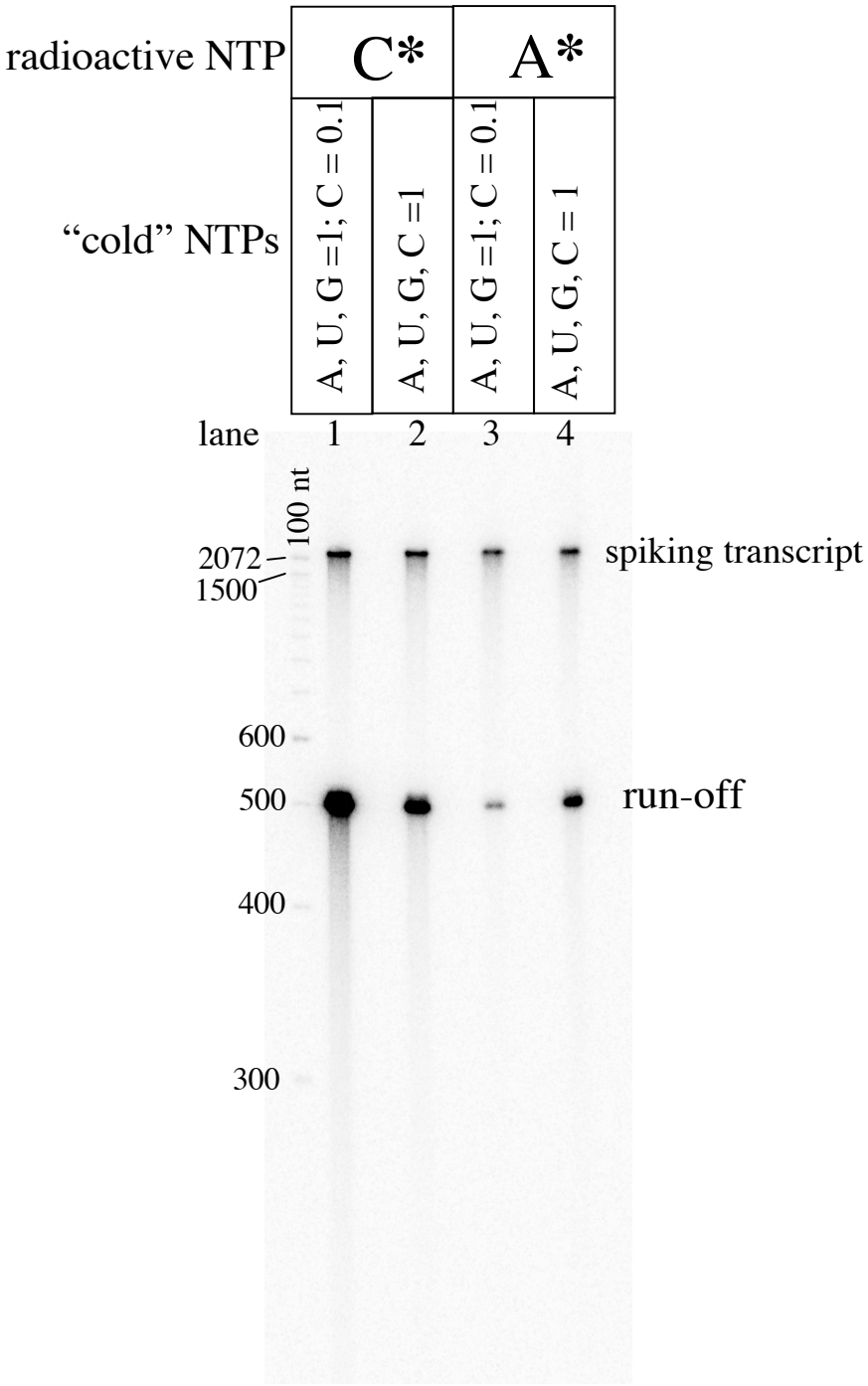
overestimate the contribution from the elongation time: if it were really contributing about 50% to the time of the transcription cycle for the longer template, from the fact that the ratio of the yields of shorter and longer transcript remains the same upon a 3-fold decrease in NTP concentration, one could conclude that the increase in elongation time should be about the same as for initiation time; otherwise, this ratio would change. However, because our measurements overestimate contribution of the elongation time to the transcription cycle, we cannot be sure that elongation time is actually detectably contributes to the time of the transcription cycle even for the longer template, thus we cannot judge how it changes with NTP concentration.

Supplementary references

1. Grabczyk, E. and Usdin, K. (2000) The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res*, **28,** 2815-2822.
2. Belotserkovskii, B.P., De Silva, E., Tornaletti, S., Wang, G., Vasquez, K.M. and Hanawalt, P.C. (2007) A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J Biol Chem*, **282,** 32433-32441.
3. Roy, D. and Lieber, M.R. (2009) G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol Cell Biol*, **29,** 3124-3133.

4. Belotserkovskii, B.P., Liu, R., Tornaletti, S., Krasilnikova, M.M., Mirkin, S.M. and Hanawalt, P.C. (2010) Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc Natl Acad Sci U S A*, **107,** 12816-12821.

Fig.S1


radioactive NTP
C*
A*
"cold" NTPs
A, U, G =1; C = 0.1
A, U, G, C =1
A, U, G =1; C = 0.1
A, U, G, C = 1
lane 1 2 3 4
100 nt
2072
1500
600
500
400
300
spiking transcript
run-off

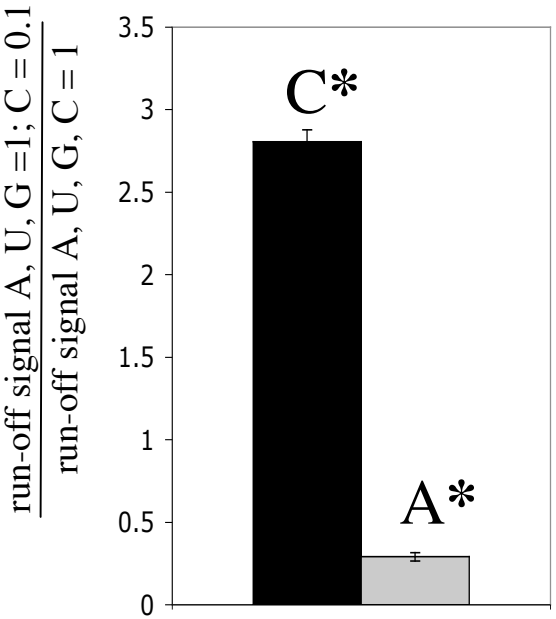

run-off signal A, U, G =1; C = 0.1 / run-off signal A, U, G, C = 1
3.5
3
2.5
2
1.5
1
0.5
0
C*
A*

Fig.S2

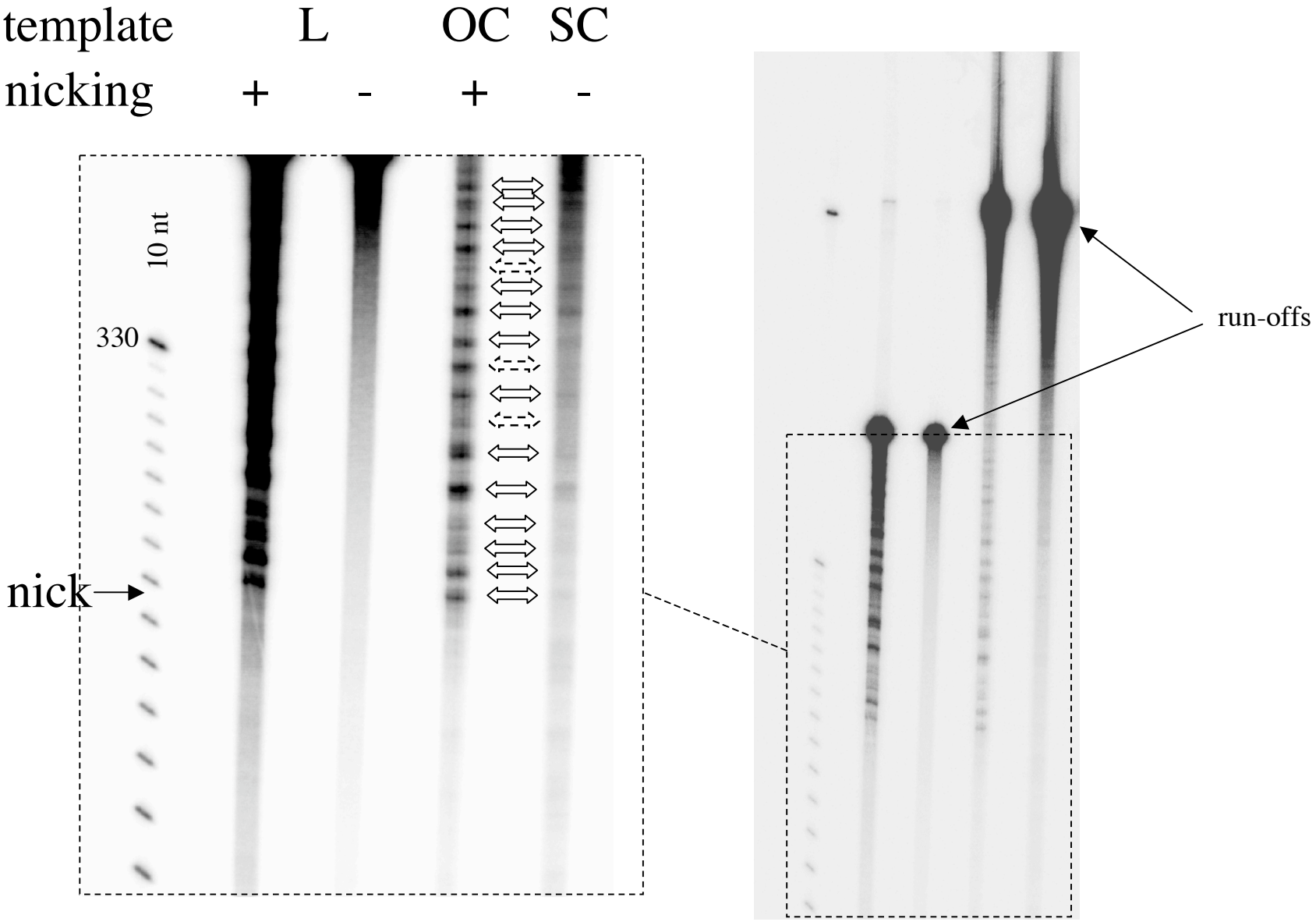


Fig.S3

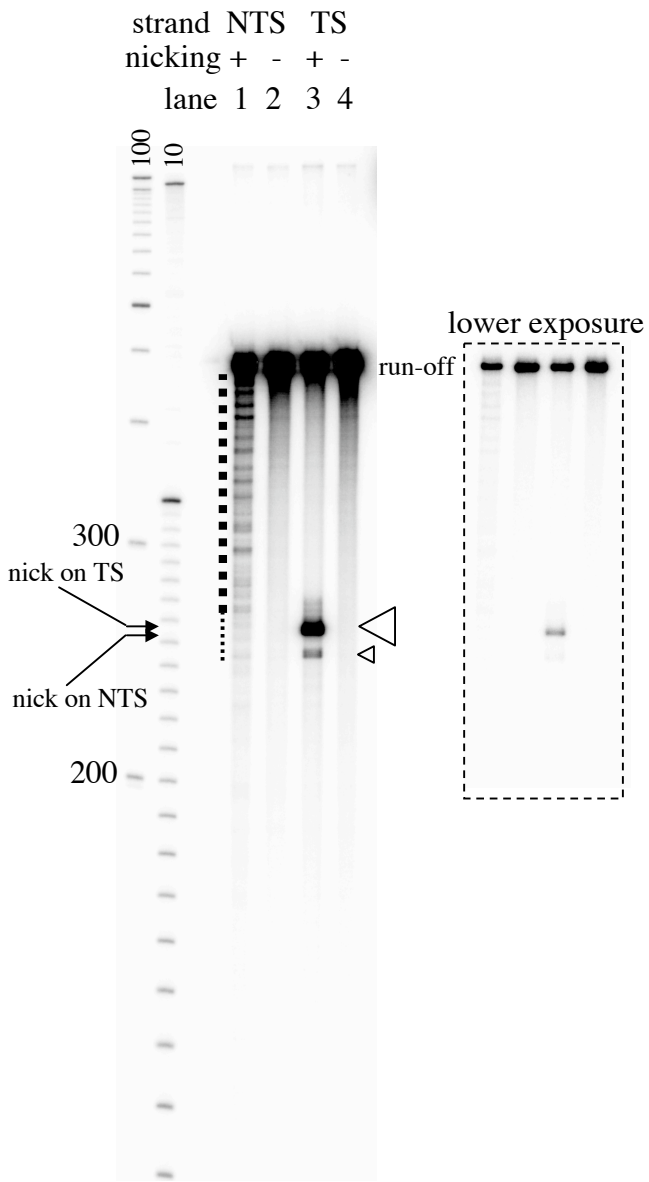

strand NTS TS
nicking + - + -
lane 1 2 3 4
100
10
lower exposure
run-off
300
nick on TS
nick on NTS
200

Fig.S4

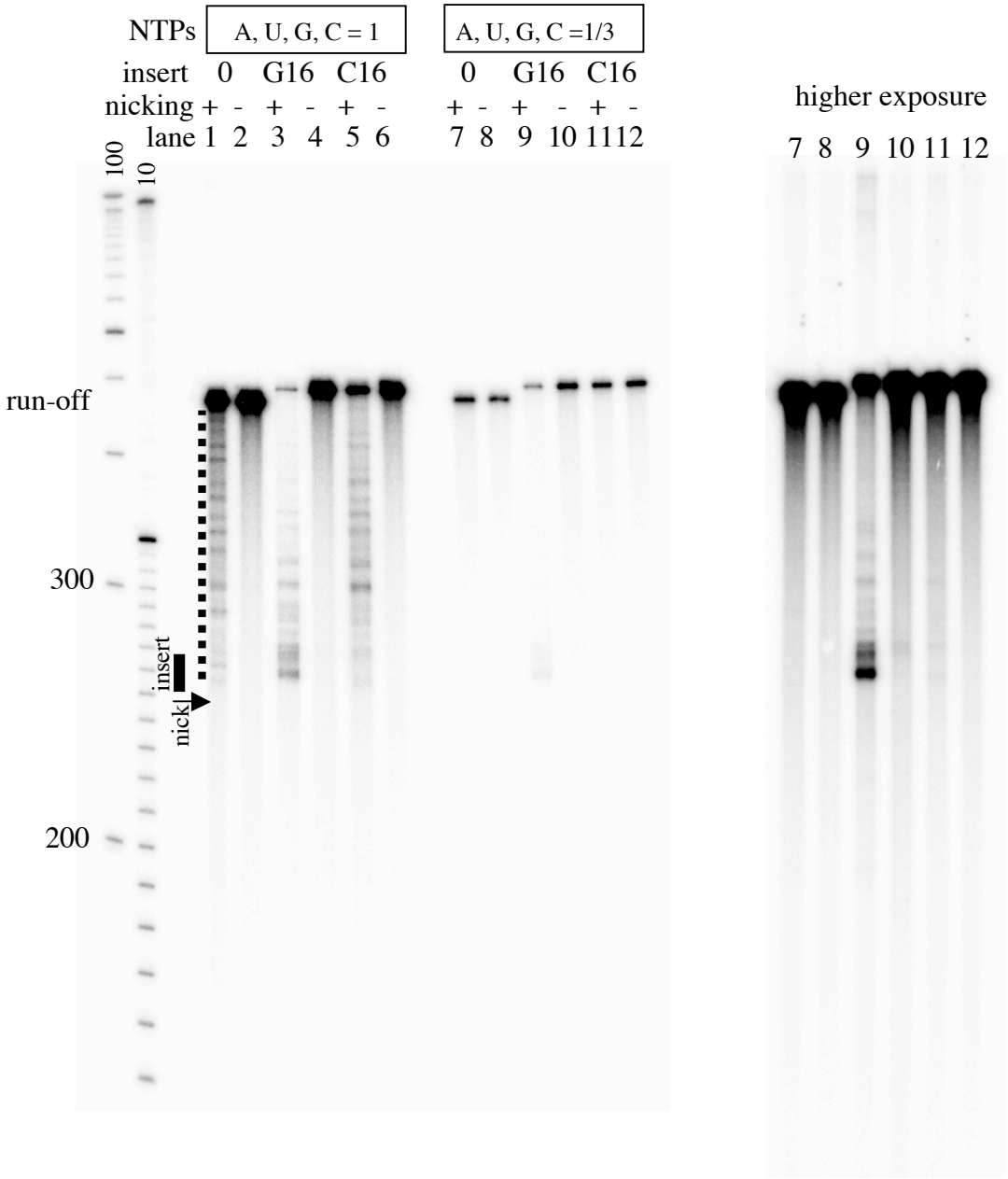

NTPs
A, U, G, C = 1
A, U, G, C =1/3
insert 0 G16 C16 0 G16 C16
nicking + - + - + - + - + - + -
lane 1 2 3 4 5 6 7 8 9 10 11 12
higher exposure
7 8 9 10 11 12
100
10
run-off
300
200
insert
nick

Fig.S5

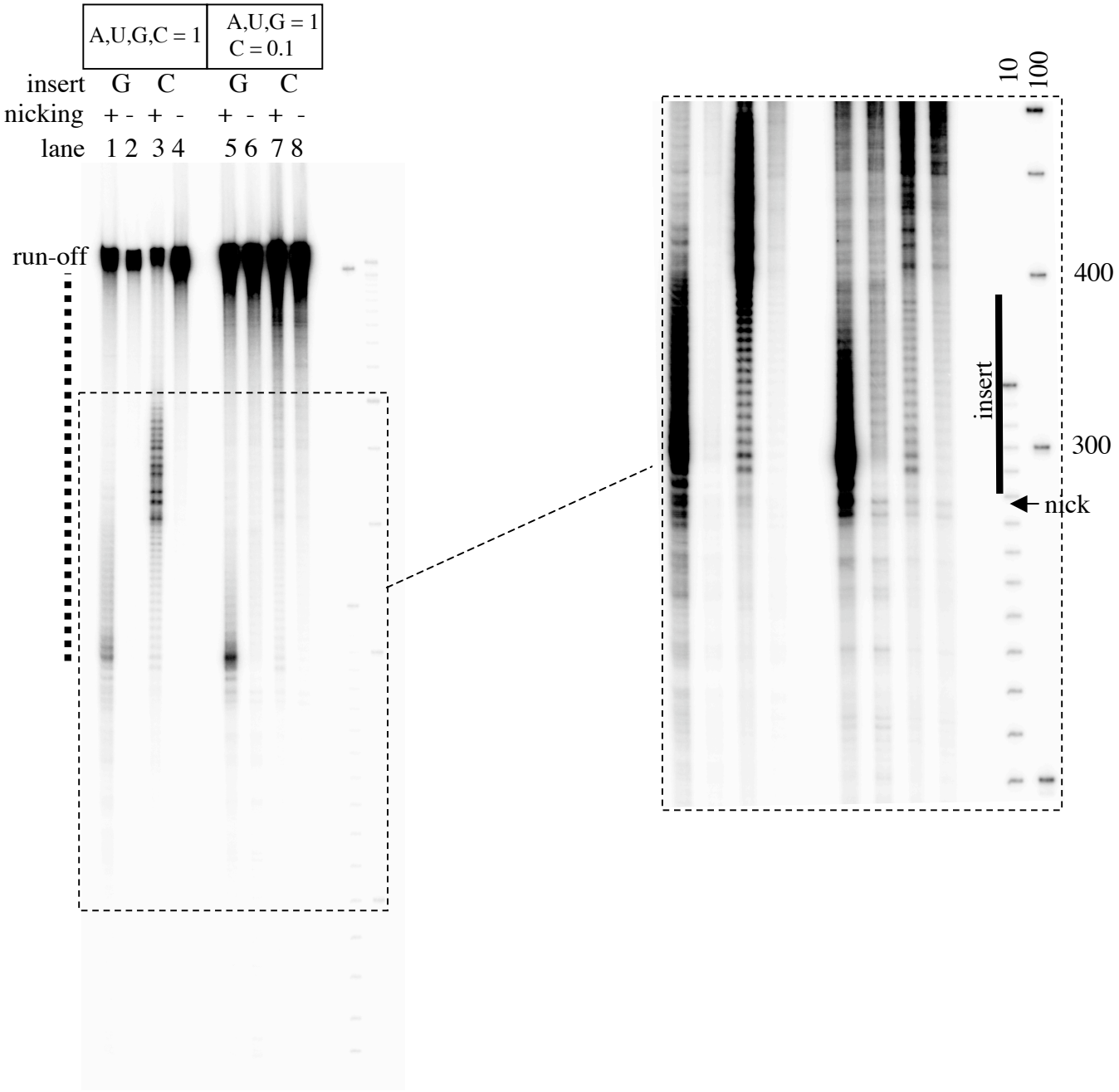

A,U,G,C = 1
A,U,G = 1
C = 0.1
insert G C G C
nicking + - + - + - + -
lane 1 2 3 4 5 6 7 8
run-off
10
100
400
300
insert
nick

Fig.S6

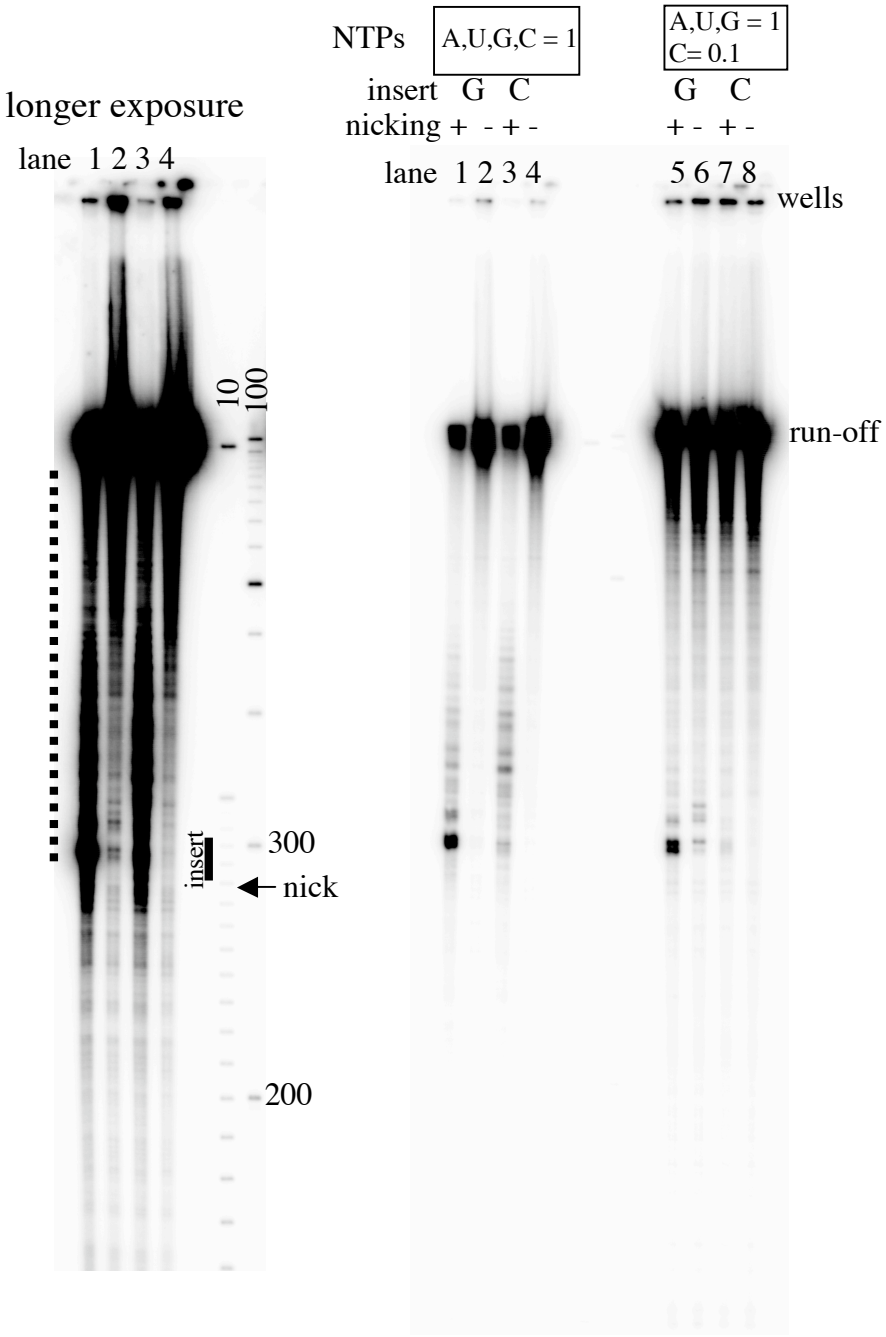

longer exposure
lane 1 2 3 4
10
100
insert
300
nick
200
NTPs
A,U,G,C = 1
A,U,G = 1
C= 0.1
insert G C G C
nicking + - + - + - + -
lane 1 2 3 4 5 6 7 8
wells
run-off

Fig.S7

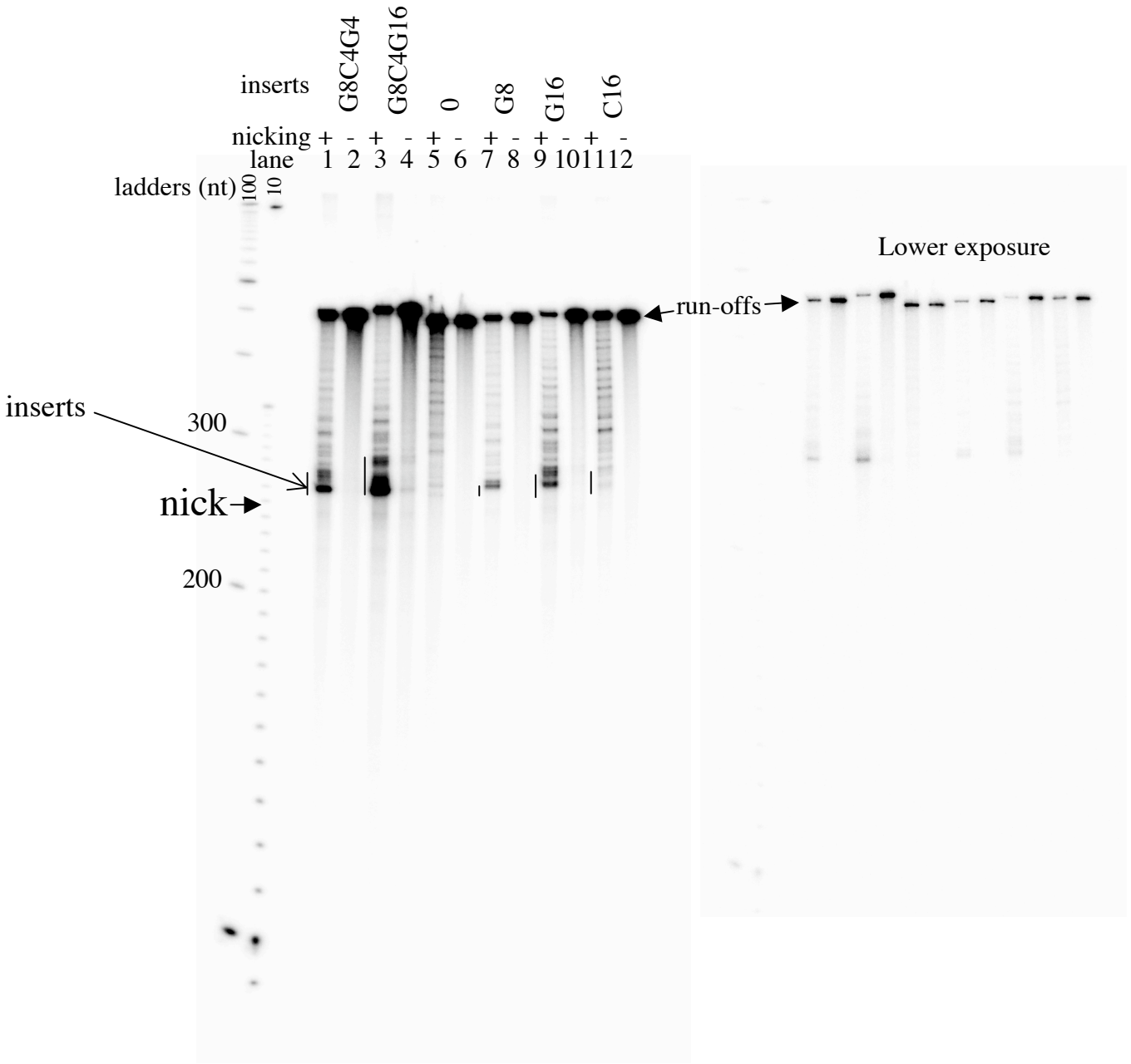

inserts
G8C4G4
G8C4G16
0
G8
G16
C16
nicking + - + - + - + - + - + -
lane 1 2 3 4 5 6 7 8 9 10 11 12
ladders (nt)
100
10
Lower exposure
run-offs
inserts
300
nick
200

Fig.S8

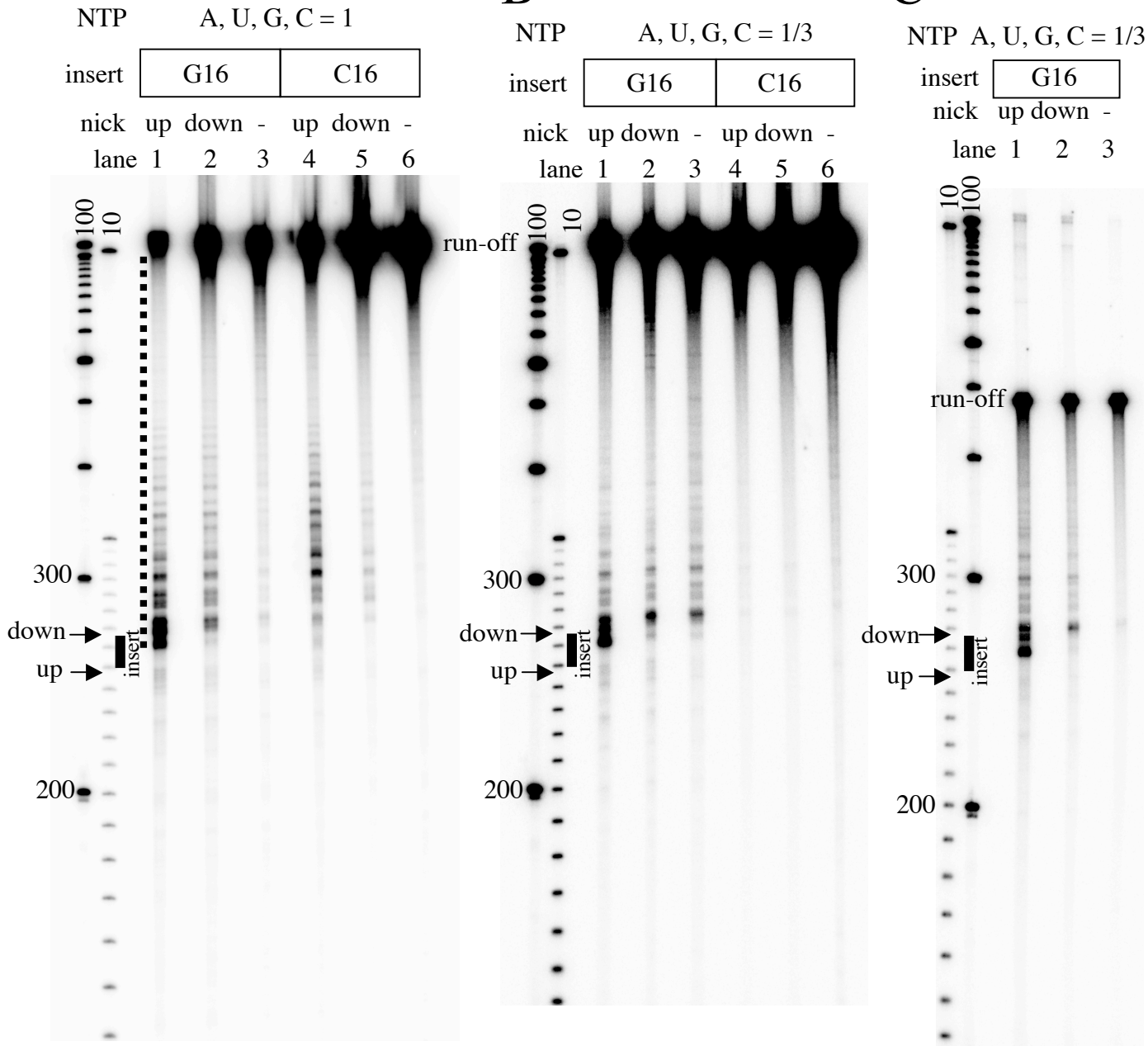

A
NTP
A, U, G, C = 1
insert
G16
C16
nick up down - up down -
lane 1 2 3 4 5 6
100
10
300
down
up
insert
200
B
NTP
A, U, G, C = 1/3
insert
G16
C16
nick up down - up down -
lane 1 2 3 4 5 6
run-off
100
10
300
down
up
insert
200
C
NTP A, U, G, C = 1/3
insert
G16
nick up down -
lane 1 2 3
10
100
run-off
300
down
up
insert
200

Fig.S9

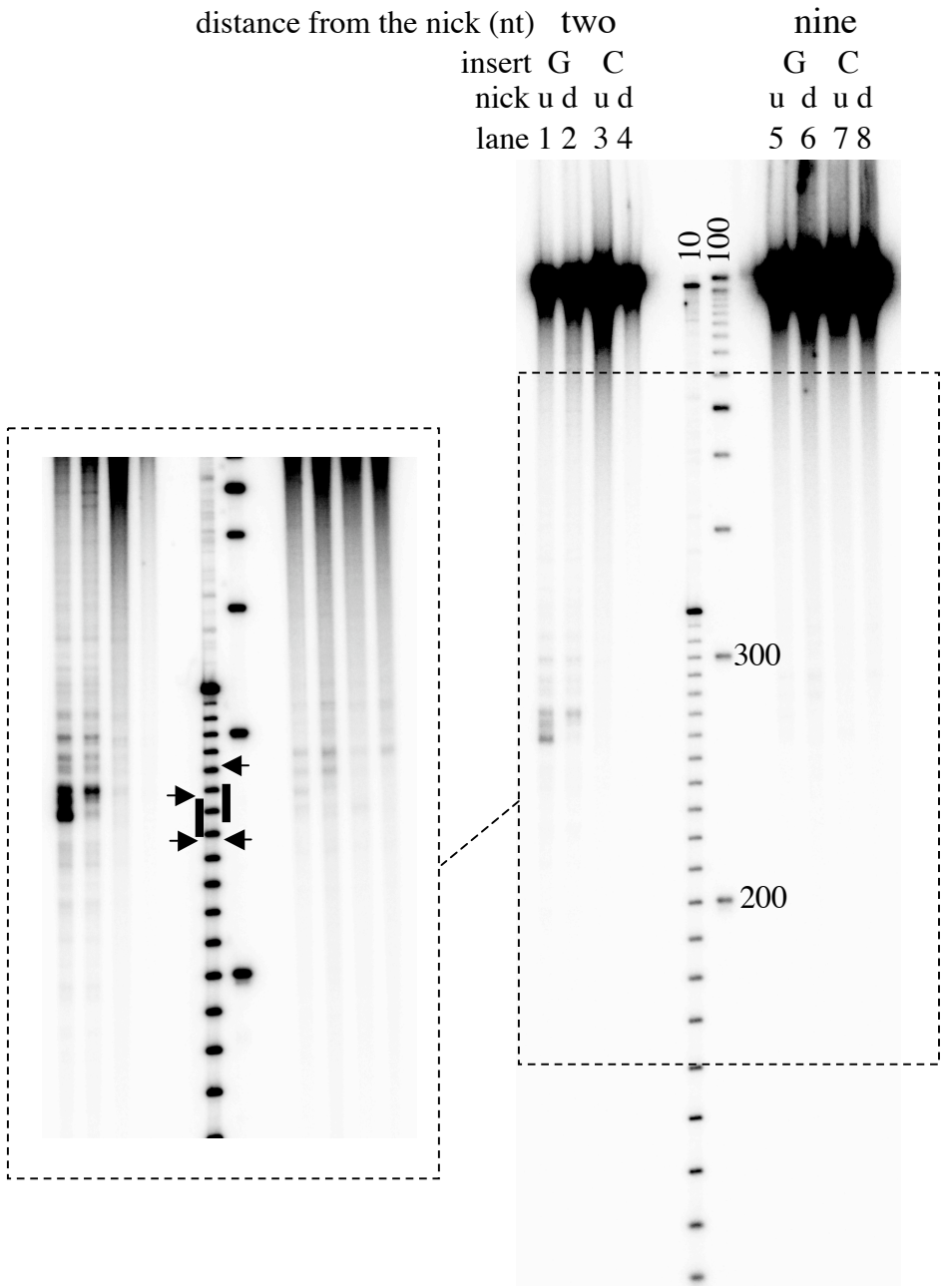
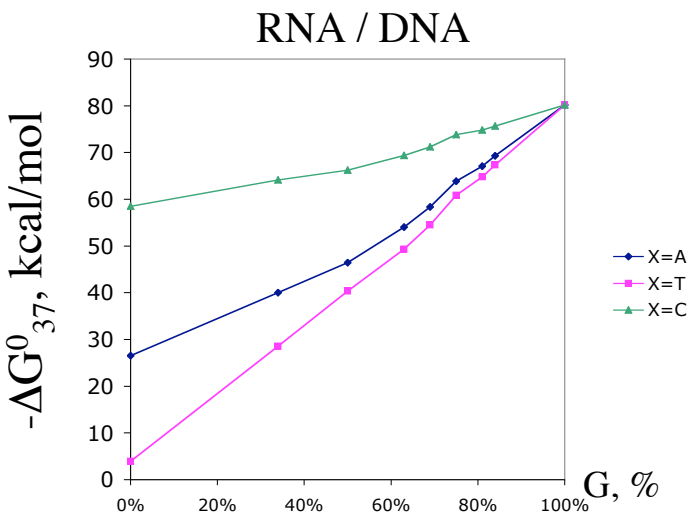
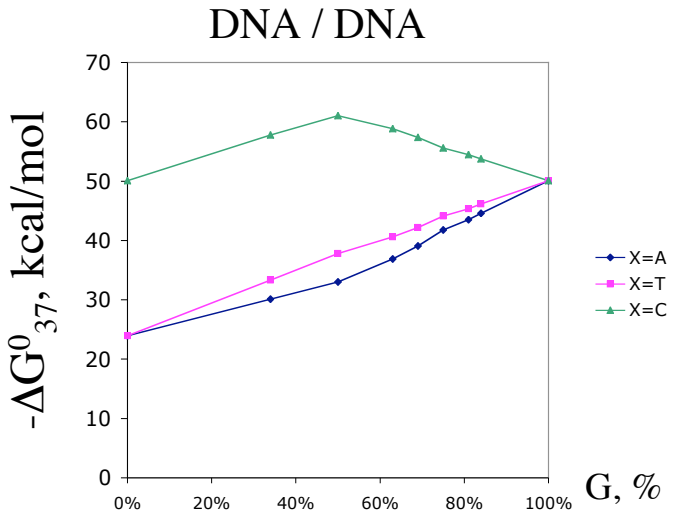

distance from the nick (nt)
two
nine
insert G C G C
nick u d u d u d u d
lane 1 2 3 4 5 6 7 8
10
100
300
200

Fig.S10


RNA / DNA
-ΔG0 37, kcal/mol
90
80
70
60
50
40
30
20
10
0
0%
20%
40%
60%
80%
100%
G, %
X=A
X=T
X=C


DNA / DNA
-ΔG0 37, kcal/mol
70
60
50
40
30
20
10
0
0%
20%
40%
60%
80%
100%
G, %
X=A
X=T
X=C

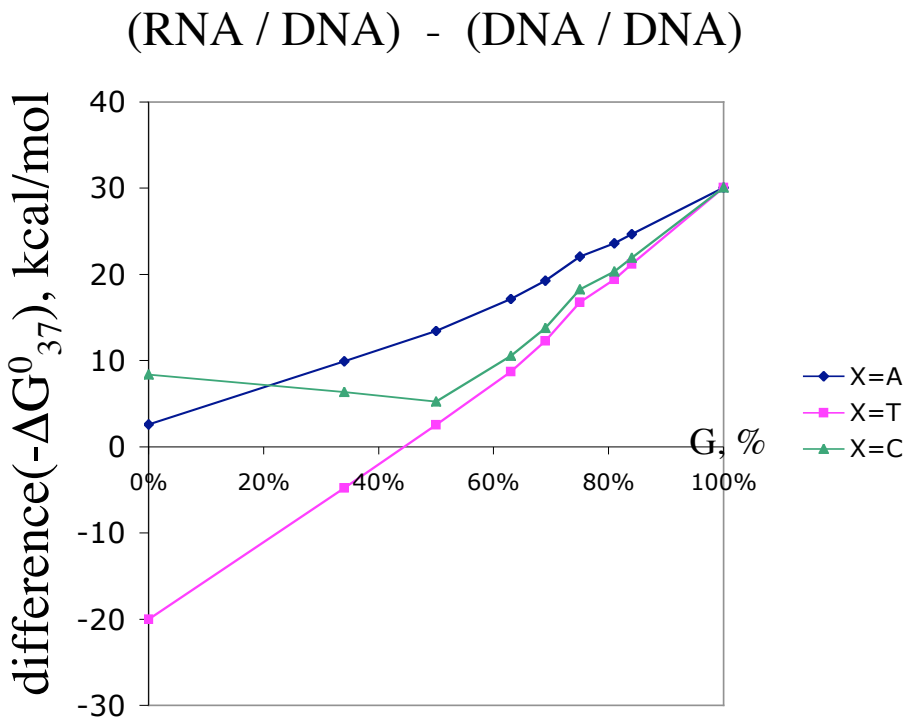
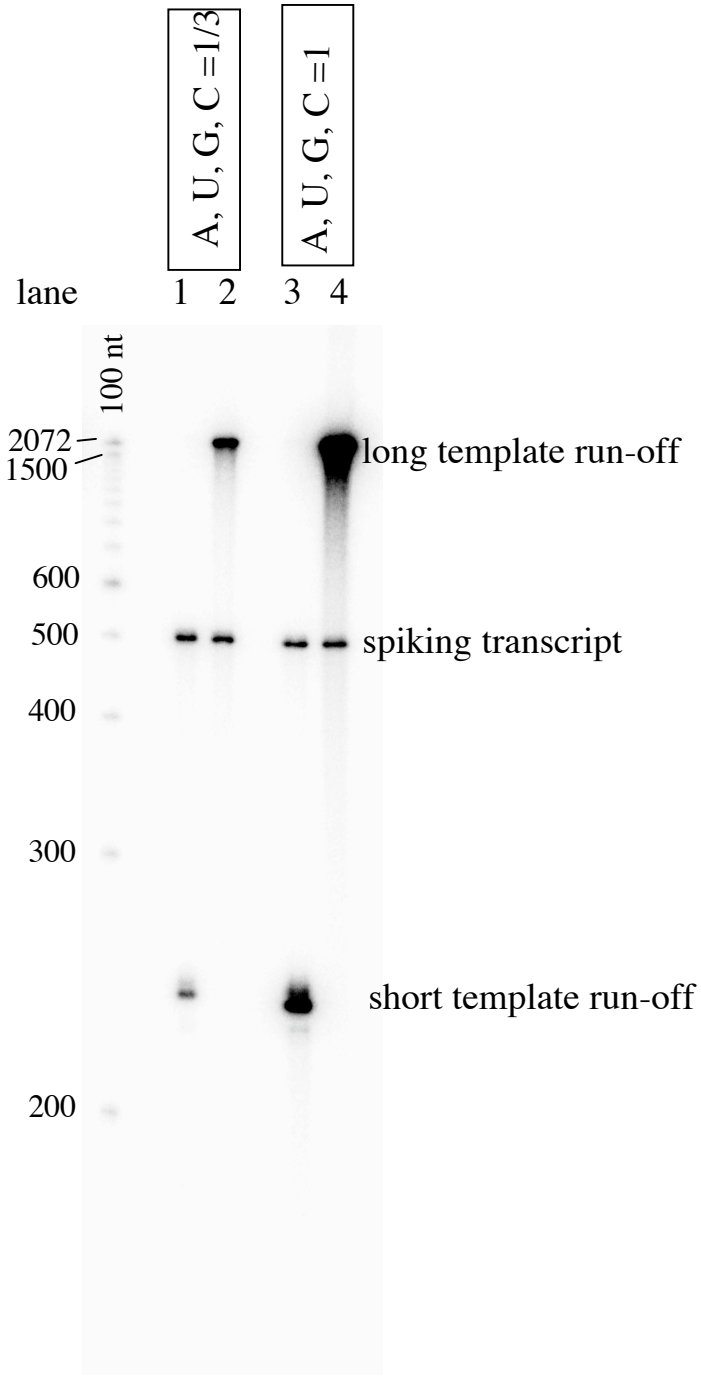
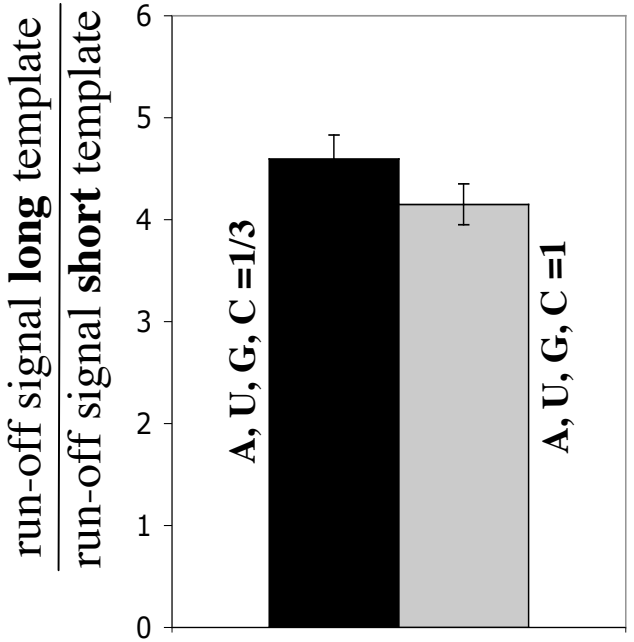

(RNA / DNA) - (DNA / DNA)
difference(-ΔG0 37), kcal/mol
40
30
20
10
0
-10
-20
-30
0%
20%
40%
60%
80%
100%
G, %
X=A
X=T
X=C

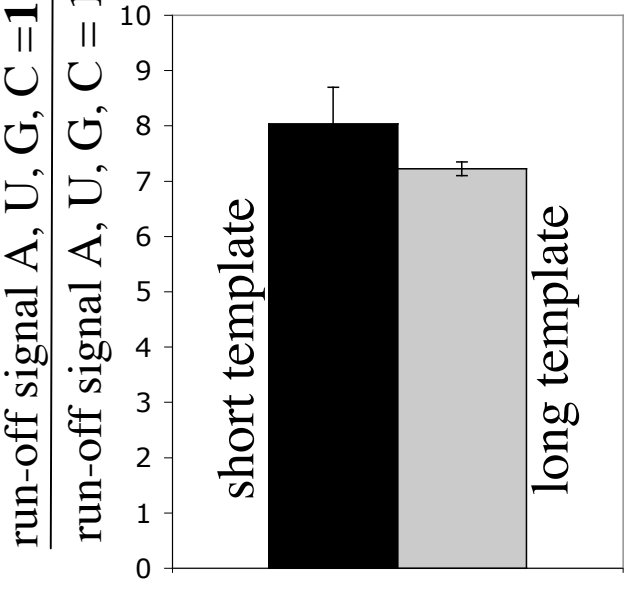
Fig.S11

TCGAGAGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGG
TTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGG
CTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTT
GGGCGCCATCTCCTTGCATGCA

Fig.S12


A, U, G, C =1/3
A, U, G, C =1
lane
1 2 3 4
100 nt
2072
1500
600
500
400
300
200
long template run-off
spiking transcript
short template run-off


run-off signal long template
run-off signal short template
0
1
2
3
4
5
6
A, U, G, C =1/3
A, U, G, C =1


run-off signal A, U, G, C =1
run-off signal A, U, G, C = 1/3
0
1
2
3
4
5
6
7
8
9
10
short template
long template